



# Rôle du temps et de la pluridisciplinarité dans la néologie sémantique en contexte scientifique. Etudes outillées en corpus

Anne Condamines, Nathalie Dehaut, Aurélie Picton

## ► To cite this version:

Anne Condamines, Nathalie Dehaut, Aurélie Picton. Rôle du temps et de la pluridisciplinarité dans la néologie sémantique en contexte scientifique. Etudes outillées en corpus. Cahiers de Lexicologie, 2012, 101, pp.161-184. halshs-00924718

**HAL Id: halshs-00924718**

**<https://shs.hal.science/halshs-00924718>**

Submitted on 7 Jan 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Introduction**

Jusqu'à il y a une quinzaine d'années, la néologie sémantique a été assez peu étudiée dans les corpus spécialisés. En revanche, elle a fait l'objet de beaucoup de travaux dans les corpus non-spécialisés. Depuis quelques années, deux éléments majeurs ont modifié la situation. D'une part, la vision d'un fonctionnement terminologique stable et parfaitement circonscrit a été remise en question et les phénomènes de variation sémantique ont commencé à être étudiés. D'autre part, le développement d'outils d'analyse de corpus a permis de tester des hypothèses sur le fonctionnement lexico-sémantique de manière systématique et rapide. Ces deux éléments ont ouvert des perspectives nouvelles dans le traitement sémantique des corpus spécialisés ; ils sous-tendent les études présentées dans ce chapitre. Un autre paramètre, également important dans ce chapitre, tient au fait que les deux études qui sont évoquées concernent des domaines scientifiques. Les corpus étudiés sont donc bien spécialisés mais la nature de leur spécialisation implique une attention particulière à la notion de nouveauté qui, dans les sciences, rencontre inévitablement la dimension extra-linguistique avec la création de nouveaux concepts.

Le chapitre décrit deux types de travaux sur la néologie sémantique en mettant l'accent sur la façon dont les outils peuvent assister la mise au jour des phénomènes en corpus scientifique. Il insiste sur la manière dont ces indices sont interprétés selon l'objectif de l'étude (le repérage de la néologie sémantique donc) et le paramètre extérieur auquel on peut associer les phénomènes de néologie, soit l'évolution dans le temps (toujours présente dans la néologie mais sur lequel nous nous focalisons dans la première étude) ou la pluridisciplinarité (deuxième étude).

La partie 1 s'intéresse aux particularités de l'étude de la néologie sémantique en contexte scientifique. La partie 2 montre comment les indices fournis par les outils sont mis en œuvre et interprétés dans chaque type de néologie sémantique.

### **1. Étude de la néologie sémantique en contexte scientifique**

Cette partie présente les enjeux des études sur la néologie sémantique en corpus scientifique et le cadre d'analyse des deux études.

#### **1.1. Néologie et terminologie**

Contrairement à la langue générale, ce n'est que récemment que la néologie sémantique a fait l'objet d'études poussées en terminologie (avec des thèses comme celles de (Assal, 1992) (Oliveira, 2009) ou (Ungureanu, 2003) notamment). En effet, la doctrine longtemps dominante en terminologie, appelée Théorie Générale de Terminologie (TGT) et édictée par le viennois E. Wüster dans les années 1930 préconisait une vision prescriptive de la terminologie. Si la néologie était considérée, c'était seulement via la création délibérée et « consciente » de termes. Cette vision ne tenait compte ni de l'évolution dans le temps, ni de l'évolution des connaissances, ni de l'évolution liée au contact avec d'autres disciplines qui

influencent souvent imperceptiblement le fonctionnement des mots. En d'autres termes, elle souhaitait protéger la terminologie de tout usage réel, associé à la « langue commune » susceptible de « pervertir » la terminologie : « En voulant effectuer le transfert inadmissible à la terminologie, des expériences accumulées dans le domaine de la langue commune, on a pendant des années opposé à la normalisation de cette dernière la devise suivante : la langue ne se laisse pas normaliser ». (Wüster, 1981 : 65).

Un tel point de vue, éminemment positiviste, s'est heurté d'une part aux besoins de traitements automatiques des documentations d'entreprises qui ont rencontré des usages de termes ou des sens non prévus et, d'autre part, aux études théoriques sur le fonctionnement du lexique qui ont montré qu'il n'était pas justifié de considérer que la terminologie aurait un fonctionnement complètement différent de celui du lexique « général ».

Dès lors, des études se sont mises en place, prenant en compte la réalité des usages dans les textes (voir ci-dessous) et la question de la néologie a commencé à être réexaminée, mais plus seulement dans une perspective de création consciente et délibérée des termes (Sablayrolles, 2008).

## **1.2. Spécificités des contextes scientifiques**

La néologie en contexte scientifique pose inévitablement la question des liens entre langue et connaissances. Il est en effet difficile d'imaginer, dans ces contextes, une évolution sémantique qui ne serait pas accompagnée d'une évolution conceptuelle et une évolution conceptuelle qui ne serait pas manifestée par une évolution linguistique. Ce deuxième point a pu être discuté. Plusieurs scientifiques ou/et philosophes des sciences pensent en effet que la création scientifique est au-delà de la dimension langagière : « La véritable création commence où finit le langage. » (Koestler, 1965). « Les mots et le langage, écrits ou parlés, ne semblent pas jouer le moindre rôle dans le mécanisme de ma pensée » (Einstein, cité par Paty, 1999).

La plupart des épistémologues admettent toutefois des relations étroites entre création scientifique et évolution linguistique, ne serait-ce que parce que la manifestation linguistique permet d'intégrer la connaissance nouvelle dans un paradigme disciplinaire. « L'invention est une séquence qui ne se limite pas aux épisodes privés de la conception et de l'élaboration, car elle n'entre dans le circuit intellectuel (et n'a une chance de jouer un rôle novateur dans la connaissance) qu'une fois communiquée » (Schlanger, 1991 : 93).

Il faut toutefois se garder de penser que ces deux types d'évolution seraient en miroir. Il existe probablement une création scientifique non verbalisable qui sous-tend la création verbalisable et il existe une création langagière qui ne concerne pas la création scientifique, même en contexte scientifique. Par ailleurs, la vision d'une néologie seulement lexicale : un concept/un terme est beaucoup trop restrictive. De la même façon qu'une langue n'est pas une suite de mots, une discipline scientifique n'est pas une suite de concepts. Comme le dit le lexicologue Guilbert : « Pour reconnaître l'emploi nouveau d'un même signe, il faut faire intervenir l'entourage du mot [...]. La relation du mot avec les autres mots de la chaîne de l'énoncé doit donc être prise en considération pour définir une typologie des néologismes. » (Guilbert, 1975 : 56).

Malgré ces difficultés à caractériser les liens entre néologie sémantique et création scientifique, nous pensons que l'étude linguistique des discours scientifiques et, tout particulièrement, l'étude de leurs contextes d'apparition peut aider à repérer de fines

évolutions linguistiques dont la mise au jour pourrait servir à éclairer des évolutions scientifiques et techniques. Cette étude est dorénavant facilitée par la linguistique outillée même si, nous allons le voir, l'aide fournie par les outils se manifeste surtout sous la forme d'indices dont la nature est en fait assez limitée et qu'il faut interpréter en fonction de l'objectif d'étude.

### **1.3. Linguistique de corpus pour l'étude de la néologie sémantique en domaine scientifique**

Les premières mises en question de la terminologie wüstérienne, dans les années 1990, ont correspondu avec le déploiement de la linguistique de corpus, ce qui a contribué à la définition d'une terminologie textuelle (Slodzian, 2000). Contrairement à l'approche wüstérienne, la terminologie textuelle préconise la prise en compte des usages réels, tels qu'ils se présentent dans les textes (au sens large, y compris les transcriptions d'oral) pour l'étude de la terminologie. Les études n'ont d'ailleurs pas toujours pour but de construire une terminologie mais, souvent, le lexique est considéré comme un moyen d'entrer dans le texte pour l'étudier dans toutes ses dimensions. Cet accès lexical aux textes est facilité par les outils qui eux-mêmes ne traitent que des chaînes de caractères, qui, par elles mêmes, n'ont aucun sens. Tout en ayant cette limite à l'esprit, on peut trouver dans ces outils une assistance à la détection de phénomènes non visibles « à l'œil nu ». Les outils peuvent fournir trois types d'indices qui, selon les corpus sur lesquels ils sont mis en œuvre et en fonction de l'objectif de l'étude (lui-même jouant un rôle majeur dans la constitution du corpus), contribueront à des interprétations différentes. Dans la linguistique de corpus ces trois indices sont le plus souvent mis en œuvre pour comparer des corpus, cette comparaison contribuant elle-même à l'interprétation. Ces indices sont détaillés dans la partie 2 mais en voici un premier panorama :

- Indices quantitatifs : dénombrement de formes et étude de leur significativité ;
- Indices de variation de formes pour un terme donné : variation concernant les monoterms (un seul mot) ou les polyterms (plusieurs mots) ;
- Indices de variation de distribution : c'est-à-dire de variation de contextes (selon le distributionalisme, le sens d'un mot, c'est l'ensemble des contextes dans lequel ce mot apparaît).

Ces indices peuvent être mis en œuvre quel que soit le phénomène linguistique étudié. Ainsi que nous le verrons, la variation sémantique est directement concernée par la variation de distribution mais la fréquence joue aussi un rôle important. La prise en compte des contextes peut se faire soit par la recherche de contextes auxquels on a associé un sens *a priori*, soit par l'interprétation de contextes fournis par des outils. Afin de mieux comprendre cet aspect, ces deux types de démarche sont détaillés dans la suite de l'article.

### **1.4. Néologie sémantique et analyse distributionnelle**

L'analyse distributionnelle, entendue au sens de prise en compte des co-textes dans lesquels apparaissent les mots, est à la base de toute analyse du sens en corpus, qu'elle soit manuelle ou automatique (Condamines, 2005). En effet, selon le distributionalisme des origines, toute modification du contexte est à associer à une modification de sens. Cette hypothèse est bien sûr particulièrement intéressante dans la perspective d'un traitement outillée des corpus. Dans une telle perspective, il faut toutefois tenir compte de différents éléments.

#### - Modifications de sens

Le changement de sens peut lui-même renvoyer à plusieurs types de phénomènes au premier rang desquels, on trouve la polysémie et la métaphore.

Dans le repérage en corpus, la polysémie est toujours associée à une variation co-textuelle, que ce soit dans une perspective manuelle (Bastuji, 1974), automatique (Bertels, 2006 ; Teubert, 2002) ou seulement outillée (Condamines et Rebeyrolle, 1996). Par ailleurs, de nombreux auteurs associent la polysémie à la néologie sémantique (Boussidan, ce numéro) : « la néologie sémantique est un cas particulier de la polysémie, avec un trait diachronique de nouveauté dans l'emploi, donc dans le sens ». (Bastuji, 1974 : 6).

Quant à la métaphore, elle est la plupart du temps associée à une modification de sens (souvent considérée comme un écart par rapport au fonctionnement attendu) qui se traduit par des modifications sémiques elles-mêmes repérables dans les contextes d'apparition (Perlerin et al., 2005).

En outre, lorsqu'elles sont étudiées dans les domaines scientifiques, les métaphores sont souvent associées à une situation de vulgarisation ou tout au moins de didactique. L'hypothèse est alors que par le recours à la métaphore, et par analogie, on peut faire comprendre à des non-spécialistes le concept qui est en jeu (Oliveira, 2009). En tout cas, la métaphore n'est pas alors considérée comme un procédé rhétorique mais bien comme un procédé « didactique ».

D'un point de vue méthodologique, il semble difficile de caractériser, sur la seule base d'une variation distributionnelle, à quel phénomène sémantique précis se rapporte le changement de contexte : polysémie ? métaphore ? néologie sémantique voire « simple » homonymie ? Dans un contexte strictement scientifique (variation temporelle ou inter-disciplinaire), il semble peu probable de repérer des fonctionnements métaphoriques si l'on entend par ce fonctionnement le recours à une analogie avec un référent plus accessible.

#### - Modification de co-texte

Dans les approches automatiques, le co-texte est entendu au sens large ; les outils utilisent de très gros volumes de données textuelles et recherchent des formes dans l'environnement immédiat des mots. Par exemple, la présence de formes similaires dans l'environnement de deux mots différents est considérée comme le signe d'une synonymie de ces mots (Bourigault et Galy, 2004). Pour les cas de métaphore, des ressources préexistantes sont utilisées (comme Wordnet dans (Mason, 2004)), qui permettent de comparer les formes « normales » donc attendues et les formes « anormales » donc potentiellement signes d'un trope. Ces approches, dont les résultats sont plus ou moins aléatoires, peuvent très difficilement être mises en œuvre avec des corpus spécialisés toujours beaucoup moins volumineux que les corpus de langue « générale ». L'étude peut être assistée mais c'est au linguiste d'interpréter les co-textes pour décider s'il perçoit une différence nette ou une similarité ou encore une simple variation. Il est évident que cette appréciation est en partie liée à l'analyse individuelle du linguiste et en partie à l'objectif de l'interprétation. C'est à la prise en compte de cet objectif et du rôle des experts que nous nous intéressons particulièrement.

#### - Rôle de la situation extra-cotextuelle

L'approche distributionnelle se limite au co-texte, c'est-à-dire, la plupart du temps, à la phrase dans laquelle apparaît un mot. Ce point de vue est restrictif car il ne tient pas compte de la cohérence globale du texte et encore moins de celle du corpus (ensemble de textes rassemblés dans le but d'une étude particulière). Or, ainsi que le pensent les tenants de la linguistique textuelle (Adam, 1999 ; Rastier, 2001), le texte constitue déjà un cadre interprétatif via les isotopies qu'il met en œuvre. Le sens d'un mot varie ainsi en fonction du co-texte immédiat mais aussi en fonction de la globalité sémantique qui sous-tend le texte. Dans cette

perspective, la position d'un mot dans le texte, son mode de mise en valeur, sa récurrence dans certaines parties du texte sont des indices qui interviennent pour comprendre son fonctionnement sémantique. Ces éléments sont très peu pris en compte par les outils de traitement automatique.

- Rôle de l'objectif d'interprétation

L'objectif d'interprétation joue, dans notre approche, un rôle déterminant tant pour la constitution du corpus d'étude que pour l'interprétation des phénomènes langagiers ou le mode d'intervention des experts. Dans les deux études que nous présentons, la dimension scientifique est également présente mais dans un cas, ce qui est étudié, c'est le rôle du temps sur l'évolution du sens et dans le second cas, c'est le rôle de la pluridisciplinarité. De ce fait, les méthodes d'analyse sont similaires mais les résultats assez différents et la perspective dans laquelle nous interrogeons les experts déterminent fortement leur démarche de validation.

- Rôle du corpus

Le corpus constitue notre objet d'étude. Il est constitué avec beaucoup d'attention pour être autant que possible lié à la problématique qui nous occupe. Nous considérons donc qu'il a sa propre cohérence et que nous pouvons nous fier aux manifestations qui s'en dégagent pour pouvoir soumettre les résultats que nous en tirons aux experts, avec une certaine assurance. Toutefois, les études que nous menons ne sont pas totalement « corpus-driven » (Tognini-Bonelli, 2001). D'une part, nous l'avons dit, elles sont orientées par l'objectif de l'étude et, d'autre part, les données sont travaillées avec nos compétences de locutrices et nos compétences de linguistes. Nous verrons dans la suite de l'article comment notre compétence *a priori* se combine avec les données du corpus pour construire des interprétations.

### **1.5. Le temps, la pluridisciplinarité : deux éléments déterminants pour la néologie sémantique en domaine scientifique.**

La néologie, qu'elle soit formelle ou sémantique, est spontanément associée au paramètre temporel : la nouveauté s'évalue à l'aune d'un avant et d'un après. Pour les domaines scientifiques, il faut convenir que la néologie peut être liée à différents types d'éléments : soit internes, soit externes à la langue. En effet, les discours scientifiques sont d'abord des discours et ils sont donc concernés par une évolution « spontanée » caractéristique du fonctionnement langagier (Nyckees, 1997). Mais ils sont aussi travaillés par la discipline elle-même qui évolue en fonction des découvertes internes et par des éléments extérieurs : demande sociétale qui oriente la recherche, apparition de nouveaux objets d'étude, événements imprévus, etc. La dimension temporelle ne prend sens que lorsque ces différents paramètres sont pris en compte. Elle peut donc produire des évolutions sur des dizaines voire sur des centaines d'années ou bien sur des durées beaucoup plus courtes. La suite de l'article rend compte d'une étude dans laquelle de courtes durées ont une pertinence, à travers l'exemple de la gestion des connaissances dans le cadre de projets aérospatiaux dits de « longue durée » (une quinzaine d'années !).

La pluridisciplinarité relève à la fois d'une évolution des disciplines (la nécessité de mieux éclairer un objet d'étude en faisant appel à d'autres approches) et d'une contrainte extérieure (les médias, les politiques, les financeurs encouragent fortement à collaborer des disciplines qui semblent travailler sur le même type d'objet). C'est le cas avec l'exobiologie dont il est question ci-dessous.

La particularité et la complexité des domaines scientifiques pour les linguistes vient de ce que, plus qu'ailleurs, les définitions scientifiques sont à la fois un lieu de reconnaissance pour les disciplines (parce qu'elles expriment leurs points de vue) et un lieu de pouvoir :

« Une définition, c'est en général la relecture d'un certain nombre d'éléments du monde à travers une théorie ; c'est donc aussi une interprétation. Ainsi, la définition d'une cellule en biologie n'est pas un point de départ mais le résultat d'un processus interprétatif théorique... ». (Fourez, 2001 : 39).

L'enjeu linguistique est donc loin d'être anodin. Que le glissement repéré par les linguistes soit lié au temps ou à la pluridisciplinarité, il peut être déstabilisant de le mettre au jour pour les disciplines concernées. Les réactions des experts sont d'ailleurs parfois assez vives et, plus qu'ailleurs sans doute, l'interprétation des résultats faite par les linguistes doit être présentée comme une proposition et élaborée en accord avec les experts. Elle peut concerner en effet l'essence même de la discipline et se heurter à des résistances assez fortes.

### **1.5.1. Cadre des études**

Les deux études présentées ont une approche commune : repérer, à l'aide d'indices fournis par des outils, des glissements de sens. Ces études prennent place au Centre National d'Études Spatiales<sup>1</sup> où l'évolution des connaissances est très rapide et où les changements interviennent dans des contextes divers. Dans cet article, nous observons les phénomènes de néologie sémantique dans deux contextes différents :

- La première étude s'intéresse à l'évolution en diachronie très courte dans le cadre de projets spatiaux dits « de longue durée » (Picton, 2009). Ces projets, qui s'étendent sur une période d'une dizaine à une vingtaine d'années, concernent par exemple l'envoi de sondes dans l'espace qui n'atteignent leur destination que plusieurs années après leur lancement. Ou encore, lorsque plusieurs générations d'un instrument sont développées, nécessitant alors un suivi de la part des experts et ingénieurs. L'objectif de cette étude est donc de mesurer en diachronie courte l'évolution des connaissances en s'appuyant sur l'exemple du domaine de l'optique et optoélectronique spatiales en diachronie très courte et sur l'exemple du projet DORIS (décrit *infra*) pour lequel un instrument, la balise de radiopositionnement DORIS, est développé depuis quatre générations.
- La seconde étude porte sur l'exobiologie, domaine qui étudie les origines de la vie sur terre et la possibilité d'une vie ailleurs. Il s'agit donc de par son objet d'étude d'un domaine particulièrement interdisciplinaire. Dans ce cas, même si un facteur temporel intervient toujours, l'objectif de l'étude est de mesurer l'impact de l'interdisciplinarité sur la constitution de la terminologie du domaine. Si la méthode utilisée est similaire, nous allons voir que l'objectif de l'étude éclaire les résultats obtenus de manière différente et les indices ne sont pas mis en œuvre de la même manière dans les deux études.

### **1.5.2. Matériau d'étude : des « corpus comparables »**

Pour saisir les glissements de sens, chacune des deux études se base sur le fait que le Traitement Automatique des Langues appelle des « corpus comparables » c'est-à-dire des corpus organisés en sous-corpus qui varient sur la base d'éléments que l'on contrôle, tels que la chronologie, le domaine (la discipline), etc. La variation linguistique observée est alors mise en relation avec ces paramètres externes de variation.

---

<sup>1</sup> Les deux études présentées dans ce chapitre ont été en partie financées par le CNES (Centre National d'Etudes Spatiales) sous la forme de bourses de thèse. Le co-financement était assuré par le CNRS (Centre National de la Recherche Scientifique).

Dans le cas de la première étude, afin de correspondre au cadre de la diachronie courte, deux corpus (eux-mêmes organisés en sous-corpus) ont été construits. Le premier est constitué des chapitres d'optique et d'optoélectronique spatiale du cours de Techniques et Technologie des Véhicules Spatiaux (désormais TTVS), édité tous les 4 ans depuis 1994 par le CNES aux Éditions Cépaduès (CNES, 1994, 1998, 2002). Ce cours est rédigé par plus de 80 experts du CNES, à l'attention de semi-experts (Bowker et Pearson, 2002) et contient une dizaine de chapitres qui englobent l'ensemble des domaines de compétence du CNES. L'attention est focalisée sur la néologie dans les domaines de l'optique et de l'optoélectronique spatiale dans cet intervalle court de 8 ans. Le corpus est constitué de 3 sous-corpus correspondant aux 3 années de parution du manuel.

Table 1. Nombre d'occurrences dans le corpus TTVS

	<b>TTVS1994</b>	<b>TTVS1998</b>	<b>TTVS2002</b>	<b>Total</b>
<b>Nombre d'occurrences</b>	46 448	78 656	109 505	234 609

Le second corpus, DORIS, est un corpus « projet » constitué des rapports de spécification des première et troisième générations de balises DORIS. DORIS (Détermination d'Orbite et Radiopositionnement Intégrés par Satellite) est un système de positionnement de satellites par balises terrestres dont la première génération a été conçue et développée à la fin des années 80 et la troisième génération au début des années 2000. Les textes sont donc répartis en deux sous-corpus : génération 1 (DORISGEN1) et génération 3 (DORISGEN3).

Table 2. Nombre d'occurrences dans le corpus DORIS

	<b>DORISGEN1</b>	<b>DORISGEN3</b>	<b>Total</b>
<b>Nombre d'occurrences</b>	17 544	18 857	36 401

Pour l'étude en contexte interdisciplinaire, le corpus est constitué de deux manuels écrits à la suite d'écoles CNRS d'exobiologie : *L'environnement de la Terre Primitive et l'Origine de la Vie* (2001), *Les traces du vivant et l'origine de la Vie* (2003). Ces ouvrages sont constitués de 38 articles de 40 auteurs différents. Ces deux ouvrages ont été divisés en disciplines, les disciplines retenues étant : l'astronomie, la biologie, la chimie et la géologie constituant 4 sous-corpus, un par discipline. Voici la répartition que nous avons obtenue :

Table 3. Nombre d'occurrences dans le corpus Exobiologie

	<b>Astronomie</b>	<b>Biologie</b>	<b>Chimie</b>	<b>Géologie</b>	<b>Corpus Exobiologie</b>
<b>Nombre d'occurrences</b>	88 815	65 589	80 190	77 010	310 000

## 2. Mise en œuvre d'indices et étude de la néologie sémantique

Cette partie consiste en une réflexion exemplifiée de la manière dont, dans l'analyse de la néologie sémantique, se met en œuvre notre compétence de linguiste, compétence qui nous permet de poser des hypothèses et de les faire évoluer en fonction des résultats proposés par les outils de traitement automatique<sup>2</sup>.

<sup>2</sup> Faute de place, nous ne présentons pas les outils que nous avons mis en œuvre. Il s'agit, pour l'étude de la néologie en contexte diachronique, de l'analyseur syntaxique Syntex (BOURIGAULT, D., FABRE, C.,



## 2.1. Indices/marqueurs ou comment se combinent hypothèses et données

La tradition en linguistique de corpus veut que l'on situe son étude comme relevant soit d'une approche « hypothesis-driven » soit d'une approche « data-driven » (Leech, 1992 ; Tognini-Bonelli, 2001 ; Rayson 2002). Ces approches se distinguent en fonction de la place donnée aux hypothèses linguistiques. Grossièrement, l'approche « hypothesis-driven » se sert du corpus pour valider des hypothèses linguistiques « introspectives ». L'approche « data-driven », quant à elle, va constituer le corpus en objet d'étude qui permettra d'établir des phénomènes linguistiques sur la base des données extraites du corpus<sup>3</sup>.

Or, cette distinction nous semble un peu trop dichotomique. En effet, si l'on prend l'exemple de la fréquence, on s'aperçoit d'emblée qu'il est difficile de la catégoriser dans l'une ou l'autre de ces approches. La fréquence, qui est l'un des indices les plus systématiquement utilisés pour l'analyse de corpus, est généralement envisagée comme une approche « data-driven » : elle est en effet le plus souvent utilisée comme un moyen d'entrer dans les corpus pour analyser par exemple les termes les plus fréquents et obtenir les résultats les plus significatifs, sans qu'il y ait d'hypothèse spécifique. C'est alors les données du corpus seules qui sont prises en compte dans l'analyse. Or, la fréquence peut être directement intégrée à une hypothèse linguistique. Par exemple, dans le cas du repérage de néologismes, le fait qu'un mot soit absent d'un corpus (donc fréquence = 0) mais fréquent dans un corpus plus récent permet une hypothèse directe sur la qualité de néologisme du mot. Dans ces cas, on adopte donc une perspective « hypothesis-driven » sur la fréquence.

Nous préférons donc nuancer cette dichotomie et envisager plutôt cette distinction comme les deux extrêmes d'un continuum pour présenter les différents indices en jeu. En effet, selon l'étude, voire selon le moment de l'étude, l'approche mettra plus en avant l'hypothèse ou les données. Dans le cas des études concernant le sens, nous favorisons une nouvelle distinction : « indice » vs. « marqueur ». La notion d'« indice » serait davantage associée à une approche « data-driven » étant donné qu'elle n'est pas liée à une hypothèse d'interprétation *a priori*. À l'inverse, le marqueur serait, quant à lui, plutôt lié à une hypothèse interprétative qui va souvent de pair avec l'objectif de l'étude. La distinction que nous faisons s'établit surtout sur le fait qu'il y a une hypothèse de base plus ou moins forte qui justifie d'emblée la constitution du corpus et son organisation en sous-corpus et qui guide en permanence l'interprétation. C'est pourquoi nous parlerons, par exemple, de « marqueur d'obsolescence ou de nouveauté » et non pas d'indice, (les marqueurs d'obsolescence étant des structures qui signalent la conscience qu'a le locuteur de la variation liée au temps, par exemple : *auparavant, on utilisait tel terme*). En effet, si nous cherchons ce type de marqueurs en corpus, c'est parce que nous posons l'hypothèse, dès le départ, que si nous trouvons des occurrences de ce type, nous serons face à une variation diachronique dans le cadre de l'étude en diachronie courte pour laquelle le corpus a été constitué. La recherche de ces marqueurs

---

FRÉROT, C., JACQUES, M.-P. AND OZDOWSKA, S. (2005) « Syntex, analyseur syntaxique de corpus ». In Actes de la 12<sup>ème</sup> conférence « Traitement Automatique des Langues Naturelles » (TALN 2005). Dourdan, France, p. 17-25.), du concordancier AntConc (ANTHONY, L. (2005) « AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom ». In Actes de la conférence "Professional Communication Conference" (IPCC 2005). Limerick, Ireland, p. 729-737.), de l'extracteur de marqueurs de relation TerminoWeb (BARRIÈRE, C. AND AGBAGO, A. , 2006, « TerminoWeb: A Software Environment for Term Study in Rich Contexts ». In Actes de la conférence internationale "Terminology, Standardisation and Technology Transfer" (TSTT 2006). Beijing, Chine, p. 103-113) et, pour la néologie en contexte pluridisciplinaire de Wordsmith, développé par Mike Scott, Oxford University, (Scott, M., 2008, *WordSmith Tools version 5*, Liverpool: Lexical Analysis Software).

<sup>3</sup> Notons que, sans lui correspondre exactement, cette dichotomie rappelle celle qui est généralement utilisée pour décrire les méthodes en sciences humaines : hypothético-déductives vs empiriques.

relève d'une approche distributionnelle puisqu'elle s'intéresse aux contextes d'apparition des termes. Mais cette recherche de contextes est sous-tendue par une hypothèse forte sur la systématique et la stabilité de l'interprétation de ces marqueurs qui ne serait pas dépendante du corpus, contrairement à d'autres marqueurs (Condamines, 2002).

À l'inverse, en ce qui concerne les indices, nous ne posons pas, dès le départ, d'hypothèse interprétative strictement liées à notre objectif d'étude. Ainsi, dans la plupart des cas d'analyse distributionnelle, il s'agit d'observer les contextes d'apparition d'un terme sans avoir d'hypothèse sur ce que l'on va trouver. Par exemple, dans le cas de la variation en diachronie, nous ne pouvons pas poser dès le départ que, s'il y a changement dans la distribution, il y a néologie sémantique. C'est justement le rôle du linguiste de caractériser la variation distributionnelle du point de vue du fonctionnement sémantique.

Pour revenir sur le cas de la fréquence, celle-ci sera toujours un indice mais selon l'étude, nous le plaçons différemment sur notre continuum « indice – marqueur ». En effet, dans le cas de l'étude en diachronie, l'indice fréquence se situe davantage du côté du marqueur tandis que, dans le cas de l'étude en interdisciplinarité, la fréquence est clairement du côté de l'indice. Il reste clair que l'observation des données s'inscrit dans une dynamique d'« aller-retour » entre hypothèse et faits observés, (c'est-à-dire, d'une part, présence/absence d'un phénomène et d'autre part conformité du fonctionnement sémantique avec ce qui était attendu). Si, une fois le corpus constitué, l'analyse peut commencer sans *a priori*, l'observation des données peut permettre de formuler une première hypothèse, puis de revenir sur cette hypothèse, de la nourrir, d'en formuler d'autres ou de repartir sur l'observation seule de données, etc. Nous illustrons ces phénomènes dans la suite de cette section.

La suite de cet article, en s'appuyant sur l'étude de ces mouvements de l'hypothèse vers les données et des données vers l'hypothèse, présente plusieurs indices et marqueurs qui nous ont permis de faire émerger différents cas d'évolution de sens. Nous les présentons en fonction du "degré" de présence d'une hypothèse linguistique dans l'exploration des données. Ainsi, nous commençons par présenter la définition de marqueurs pour repérer des changements sémantiques, puis, dans un second temps, nous décrivons l'analyse de la distribution de termes. Dans le cas de l'analyse diachronique, nous verrons que la description de la distribution est conditionnée par une hypothèse sur les variations de fréquence des unités linguistiques structures comportant un terme alors que dans le cas du contexte de pluridisciplinarité, la distribution est au contraire envisagée comme un indice à part entière, et se rapproche donc d'une perspective « data-driven » (ou proche des indices dans la vision de continuum proposée).

## **2.2. Analyses du changement sémantique en corpus**

Cette partie montre, à l'aide d'exemples, comment se met en œuvre le processus interprétatif des phénomènes de néologie sémantique, à partir des indices/marqueurs fournis par les outils. Nous examinons trois types de cas.

### **2.2.1. Changements sémantiques repérés à l'aide de marqueurs**

Nous présentons ici un cas où la recherche de marqueurs est directement liée à l'objectif de l'étude. Dans le cas de l'étude en interdisciplinarité, il s'agit des marqueurs qui montrent la conscience de l'auteur d'une possible polysémie du terme qu'il utilise. Des expressions comme « au sens de [telle discipline] » ou encore « en [telle discipline] » ont ainsi été recherchées dans le corpus à l'aide d'un concordancier.

ex. 1 : Il y a peu de concepts en chimie physique macroscopique qui soient aussi abstraits que le concept d'entropie ; il y a peu de concepts qui soient si souvent mal utilisés, tout particulièrement en biologie et de manière plus générale dès qu'il s'agit de décrire des processus irréversibles conduisant à des états plus structurés que l'état de départ.

ex. 2 : De nombreux auteurs ont émis l'hypothèse que la tectonique des plaques ne fonctionnait pas pendant l'Archéen et donc il a même été proposé que, compte tenu des forts flux thermiques, il n'y ait pas eu de lithosphère archéenne (dans le sens rhéologique du terme).

Ces exemples illustrent des cas de la migration d'un terme d'un domaine à un autre. L'exemple 1 illustre le fait que l'emprunt d'un terme à une discipline entraîne un jugement de valeur de la discipline source envers les autres disciplines qui emploient un de leurs termes. En effet, les chimistes perçoivent le fait que l'utilisation du terme *entropie* par des biologistes peut entraîner une certaine « déformation » du sens parce qu'il y a une méconnaissance de la notion. Dans ce cas précis, pouvons-nous considérer que le sens qui va s'installer au fil du temps chez les biologistes est un cas de néologie sémantique ?

Un autre exemple du même type, relatif à un emprunt terminologique, est celui d'un terme qui est emprunté à une discipline mais qui existe déjà dans la discipline receveuse, avec un sens différent. Cet état de choses peut entraîner des situations d'ambiguïté où l'on se demande quels sens l'on doit donner à un terme. Pour illustrer cette situation nous traiterons du cas de *réduction* et de ses dérivés dans le corpus d'exobiologie. Le terme *réduction* est emprunté par les autres disciplines à la chimie, emprunt complètement assumé puisqu'on retrouve des marques de cet emprunt comme illustré dans l'exemple suivant :

Ex.3 : Du point de vue chimique, une oxydation est définie comme la perte d'un ou plusieurs électrons par un atome ou une molécule. Une réduction est, elle, définie comme l'addition d'un ou plusieurs électrons à un atome ou une molécule. (Biologie)

Dans les cas où l'emprunt n'est pas explicite, on réussit à associer *réduction* ou ses dérivés à sa définition chimique grâce au contexte qui souvent fait référence aux phénomènes d'oxydation et de réduction. Comme dans les exemples suivants :

Ex.4 : <equation> est une réaction d'oxydation, qui s'accompagne en présence d'O<sub>2</sub>, d'une réaction de réduction : <equation>. (Biologie)

Ex. 5 : L'oxydation d'une Terre réductrice est alors nécessaire, ne serait-ce que pour permettre la ségrégation du métal (...). (Géologie)

En revanche, on trouve également des occurrences de l'utilisation de *réduction* dans son sens général, sens qu'on ne retrouve pas dans le sous-corpus « chimie » :

Ex.6 : Les mécanismes d'évolution par simplification et réduction de la taille ont joué un grand rôle dans l'histoire du vivant terrestre. (Biologie)

Ex.7: Dans tous les cas, on observe une réduction de la taille du génome (en bleu). (Biologie)

En biologie, le sens général de *réduction* et le sens spécialisé de la chimie vont cohabiter et pourrait occasionner certaines ambiguïtés. Comme dans les exemples suivant où le sens à attribuer à *réduction* est moins clair :

Ex.8 : (...) il s'agissait d'anciens champignons qui se sont simplifiés par réduction, suite, encore une fois, à un mode de vie parasitaire. (Biologie)

Ex.9 : La RNR ancestrale était déjà très certainement une protéine, car le mécanisme de réduction du ribose fait intervenir des groupements radicalaires très réactifs qui n'auraient sans doute pas pu exister de façon stable dans un ARN. (Biologie)

D'une certaine façon, l'emprunt du mot *réduction* à la chimie conduit à une spécification du sens initial dans ce domaine (et celui qui est communément admis), ce qui peut créer une ambiguïté. La question peut se poser alors de savoir si on n'a pas à faire à une néologie sémantique dans la discipline receveuse.

### 2.2.2. Changements sémantiques repérés à l'aide de la distribution associée à des hypothèses sur les variations de fréquence *a priori*

Le type d'analyse que nous présentons dans cette section repose sur l'étude de la distribution et sur un *a priori* moins net que dans le cas précédent quant à l'indice observé. Cependant, l'observation est somme toute orientée sur la base d'une hypothèse liée à la variation de fréquence des distributions observées. Les exemples que nous proposons sont tirés de l'analyse diachronique en optique spatiale. En effet, tout particulièrement dans l'étude diachronique, nous faisons l'hypothèse que, pour observer un changement sémantique, non seulement les distributions du terme analysé doivent changer entre les différents sous-corpus, mais la fréquence de ces modifications doit aussi changer, c'est-à-dire qu'elles peuvent apparaître/disparaître ou significativement croître/diminuer.

Le cas de *parasite*

Si l'on observe la distribution de la forme *parasite*, voici les données que l'on peut obtenir dans les trois sous-corpus du TTVS. Ces données concernent les cas où *parasite* apparaît en position de modifieur (ou épithète) d'un nom (créant alors ce que l'on appelle un *descendant* (Bourigault, 2007)).

Table 4. Exemple de changement de distribution – le cas de *parasite* (TTVS)

Terme (modifieur)	Descendants	Fréquence		
		TTVS1994	TTVS1998	TTVS2002
<i>parasite</i>		40	87	142
	<i>lumière parasite</i>	18	34	66
	<i>flux parasite</i>	3	16	8
	<i>source parasite</i>	3	5	3
	<i>élément parasite</i>	2	5	6
	<i>signal parasite</i>	0	3	5
	<i>rayonnement parasite</i>	0	0	4
	<i>image parasite</i>	0	7	17
	<i>flux lumineux parasite</i>	0	4	2
	<i>trajet parasite</i>	0	2	5

Ce tableau illustre l'apparition, au fil du temps, de nouvelles combinaisons qui sont plus spécifiques que celles déjà existantes en 1994. En effet, *flux*, *source*, *lumière* et *éléments parasites* sont complétés par de nouveaux éléments parasites, plus spécifiques, tels que *image*, *trajet* ou *signal* qui n'apparaissent qu'en 2002. Il semble alors qu'avec le temps, les locuteurs ont recours à des termes/concepts beaucoup plus diversifiés pour appréhender la notion de *parasite*.

L'analyse peut cependant être complétée par une observation en corpus (et donc sans filtre de fréquence). On s'aperçoit alors rapidement que les verbes dont l'un des arguments est un SNom qui contient la forme *parasite* sont eux aussi grandement diversifiés et précisés. En effet, en 1994, les contextes verbaux contenant *parasite* sont rares et se limitent à décrire le fait que les éléments parasites sont indésirables, qu'il faut les extraire ou les éviter :

Ex.10 ... de **rejeter** le maximum de lumière parasite, ...  
... chaque détecteur **ne doit pas recevoir** de flux parasite...

En 2002, ces contextes sont beaucoup plus nombreux et variés et montrent un regard nouveau et plus « complet » sur cette question. En effet, on entrevoit le développement de nouvelles solutions marquées par l'emploi de verbes tels que *permettre* ou *arriver à* (exemple 11) qui sont utilisés pour expliquer les nouvelles solutions existantes.

Ex.11 ... permet de **s'affranchir** des parasites. Des essais sur l'instrument complet permettent...  
... reflets permettent également **d'éliminer** les images parasites qui peuvent se ...  
... permet de **réduire** les réflexions parasites par une géométrie soignée...  
... utile, de façon également à **limiter** le flux parasite. Le refroidissement...  
... pour **arriver à** une bonne **suppression** de cette lumière parasite...

D'autres contextes permettent d'affiner l'explication de cette évolution dans la mesure où ils mettent en évidence l'existence de nouvelles possibilités liées à de nouveaux moyens de modélisation informatisée de ces parasites (exemple 12) :

Ex.12 ... **Modéliser** la lumière parasite dans un instrument requiert des logiciels...  
... On **identifiera** les trajets parasites par **modélisation**...

Il est donc clair qu'en 2002, il ne s'agit plus d'avoir conscience de ce phénomène indésirable, mais d'apporter des solutions fines et efficaces pour lutter contre eux et les éliminer.

Enfin, si l'on analyse le terme *lumière parasite* qui est le syntagme le plus fréquemment formé avec la forme *parasite*, on observe le même type de changements contextuels, résumés dans le Tableau 5 ci-dessous.

Table 5. Changement de distribution – *lumière parasite* (TTVS)

TTVS1994	TTVS2002
filtrage de ~ piégeage de ~ suppression de ~	filtrage de ~ piégeage de ~ suppression de ~ <b>lutter</b> contre ~ <b>réduire</b> ~ <b>contrôle</b> de ~ <b>spécification</b> de ~ <b>étude</b> de ~

Ce type de cas relève de ce que l'on pourrait qualifier de *mise à jour* du sens, ou, plus finement dans ce cas, de *précision du sens* : les connaissances avançant, il arrive que le stock de connaissances qui existent pour un concept ou un phénomène donné augmente. Dans ce cas, la définition du terme/concept doit être complétée au fur et à mesure que l'on sait mieux le saisir et le connaître.

Dans cet exemple, il serait certainement trop hâtif de parler de néologie sémantique, mais il est clair que les connaissances des experts sur le terme/concept *parasite* ont augmenté et que lorsque cette question est abordée en 2002, elle l'est de manière plus précise qu'en 1994. En d'autres termes, lorsque le locuteur aborde la question des parasites en 2002, il le fait en tenant compte des connaissances actuelles des interlocuteurs sur cette question, de leur formation, etc., qui sont différentes de celles de 1994. Dans cet exemple, le changement de distribution ne donne pas directement accès aux nouveaux traits sémantiques à prendre en compte, mais ils permettent de mettre en avant une utilisation nouvelle de ce terme/concept : si celui-ci reste le même « objet », on doit prendre en compte les progrès dans les connaissances dont les experts disposent sur ce terme/concept dans sa définition.

#### Le cas d'*opérateur*

À partir de l'analyse de la distribution et de la fréquence, ce type d'interprétation n'est cependant pas la seule possible. Dans le corpus DORIS cette fois-ci, un autre cas de figure peut être illustré par l'exemple d'*opérateur*, terme pour lequel on observe un changement de contextes d'utilisation marqué. Dans cet exemple, le phénomène de changement que nous illustrons relève plus d'un changement de *point de vue*, qui peut également être considérée comme un aspect différent de la notion de *mise à jour* introduite *supra*.

Table 6. Exemple de changement de point de vue – le cas d'*opérateur* (DORIS)

		Fréquence	
Expansion	Descendants	DORISGen1	DORISGen3
<i>opérateur</i>		89	26
	<i>interface opérateur</i>	0	12
	<i>opérateur appuyer sur touche</i>	26	0
	<i>opérateur appuyer sur val</i>	5	0
	<i>opérateur appuyer</i>	33	0
	<i>opérateur tourner clé</i>	6	0

La notion de « points de vue », comprise comme autant de façons de concevoir et verbaliser la réalité, a été notamment abordée en langues de spécialité dans Condamines et Rebeyrolle (1996). L'étude d'un corpus du spatial a permis d'identifier six emplois différents du terme *satellite* (c'est-à-dire six catégories de contextes d'utilisation), emplois qui ont pu être associés à autant de « points de vue » : « corps artificiel », « objet mobile », « plate-forme », « véhicule », « hôte », « relais ».

Ces six points de vue ont été identifiés à partir des structures lexico-syntaxiques dans lesquelles il est construit en discours, et ne reflètent pas réellement un fonctionnement polysémique, mais plutôt différents points de vue sur cet objet : « Ce que nous retiendrons des [six] types de constructions mis en évidence c'est que le terme *satellite* n'est pas

véritablement polysémique mais plutôt que coexistent différentes façons de l'aborder, différents angles depuis lesquels on peut considérer cet objet [...]. Il serait [...] erroné de dire que le terme *satellite* est monosémique, néanmoins, bien qu'il ne s'agisse pas non plus de polysémie, le mot reste aussi ambigu que dans la langue générale. » (Condamines et Rebeyrolle, 1996 : 49).

Cette notion de *point de vue* est intéressante pour l'étude de la néologie sémantique dans la mesure où elle s'inscrit dans une perspective discursive : les différents points de vue sur un terme donné sont observables en discours, à travers notamment la combinatoire des termes dans les textes produits.

Dans l'exemple que nous présentons, la nouvelle structure *interface opérateur* renvoie à un élément nouveau sur la balise DORIS. Mais au-delà de cet aspect, il est intéressant d'observer la disparition massive de toutes les dépendances verbales où le verbe implique une action manuelle dont l'opérateur est l'agent. Cet élément laisse entrevoir le changement du rôle de l'opérateur sur la balise DORIS : il disparaît en tant qu'agent d'une action sur la balise (*appuyer, tourner la clé*) mais apparaît associé au terme *interface*. Le néologisme *interface-opérateur* révèle l'influence de l'informatique dans le fonctionnement des balises et sur le rôle de l'opérateur. Ce dernier a grandement évolué en dix ans et l'on peut aisément émettre l'hypothèse que la formation et les responsabilités de l'opérateur DORIS ne sont plus les mêmes aujourd'hui. Cette analyse illustre bien la complémentarité de l'observation des dépendances syntaxiques et de leurs variations de fréquence dans une perspective diachronique.

Cependant, encore une fois, bien que le changement sémantique associé au terme/concept *opérateur* soit sensible, peut-on parler pleinement de néologie sémantique ? L'opérateur reste l'agent humain qui intervient sur la balise, mais son rôle et ses « moyens d'intervention » sont différents en 2000 (génération 3) par rapport à 1989 (génération 1). Il s'agit donc du même sens associé au terme/concept, mais dont l'évolution doit être prise en compte par exemple pour la maintenance des ressources en tant que mise à jour du sens repérable en corpus.

Dans les deux cas de figure présentés, nous montrons la manière dont l'analyse de la distribution, guidée par un filtre de fréquence, permet de mettre au jour différents aspects de l'évolution sémantique d'un terme. Si dans les deux cas le changement peut être associé à la notion de « mise à jour » une différence importante doit être considérée entre ces deux exemples :

- dans le premier cas, le changement sémantique est lié à une meilleure connaissance des propriétés d'un terme/concept que l'on doit prendre en compte pour maintenir à jour sa définition. Le sens est précisé dans le temps : le concept *parasite* n'est pas modifié. Les experts découvrent progressivement les propriétés et la nature de ces parasites, mais le concept en lui-même ne change *a priori* pas ;
- dans le second cas, le changement sémantique est lié à un changement de point de vue sur le terme/concept considéré. Le terme/concept n'est plus envisagé de la même manière : le concept *opérateur* est modifié, ne serait-ce qu'au niveau des tâches que cet agent accomplit sur la balise aujourd'hui par rapport à 1990.

Si cette distinction ne remet pas en cause l'évolution observée, il peut être important de la prendre en compte pour décider si elle implique l'apparition d'un nouveau sens ou non. En effet, la notion de *point de vue* permet d'adopter une vision continue sur le changement sémantique dans le temps, entre monosémie et polysémie. Sur ce continuum, on peut basculer du changement de point de vue vers l'émergence d'un nouveau sens dans la langue.

Identifier, ou peut-être plutôt décider, l'emplacement de ce point n'est pas chose aisée et relève très certainement de l'interprétation de l'analyste et des experts, qui devront décider si ces indices distributionnels manifestent l'apparition d'un nouveau sens ou d'une mise à jour à opérer dans la définition de ce même objet. Dans la mesure où ces types d'évolutions sont observables en discours, ceci implique que les évolutions en question sont associables à différents groupes de locuteurs identifiables, que les discours varient en fonction de ces locuteurs, de leurs relations et des situations de communication (Beacco et Moirand, 1995 : 34). Transposé sur la dimension diachronique, cela signifie que les discours sont adaptés dans le temps, en fonction des locuteurs impliqués dans le discours, de la période à laquelle ils sont produits, et surtout des connaissances dont les locuteurs disposent sur les termes/concepts impliqués dans ces discours. Et ce sont ces connaissances, repérables en corpus, qui peuvent être prises en compte pour maintenir et gérer la définition des termes/concepts.

### 2.2.3. Changements sémantiques repérés à l'aide de la distribution associée à des interprétations sur les variations de fréquence *a posteriori*

Le type d'analyse que nous présentons dans cette section diffère de celle présentée dans la partie précédente en ce qui concerne la place donnée à la variation de fréquence. En effet, ici, il s'agit d'étudier la distribution d'un terme sans attente et c'est seulement une fois qu'un phénomène intéressant est détecté que sa fréquence est recherchée.

Le cas d'*accrétion*

L'environnement distributionnel du terme *accrétion* a été observé dans le cadre de l'étude en interdisciplinarité. Ce terme a été choisi parce qu'il était présent dans nos quatre sous-corpus et qu'une définition en était donnée dans le glossaire de l'ouvrage. Il s'agissait donc de regarder son fonctionnement dans chaque sous-corpus. Premier constat donc, les différentes disciplines concernées par l'exobiologie (la biologie, la chimie, l'astronomie et la géologie) utilisent le terme d'*accrétion*. Ce terme est généralement utilisé en astronomie pour illustrer le processus d'agglomération de matière conduisant à la formation de corps de taille supérieure à l'objet d'origine (Gargaud, 2001). En recherchant ce terme à l'aide d'un concordancier afin d'observer son environnement distributionnel, on remarque qu'il est utilisé généralement, en astronomie, comme un processus dans des structures comme *durant/lors de/ au début de/ en fin de + l'accrétion*. Cependant, dans certains cas, le terme ne renvoie plus au processus mais à ce que l'on peut considérer comme le résultat du processus, avec des syntagmes tels que *disque d'accrétion, accrétion de planétésimaux/ de planètes/ de matière/ de roche, glace d'accrétion*. Or, ce second usage est bien plus présent en biologie, chimie et géologie.

*Accrétion* comme processus :

Ex.13 : – fort ralentissement de l'accrétion des solides (cf. ci-dessus) et *accrétion lente de gaz de la nébuleuse, à un taux à peu près constant pendant plusieurs millions d'années, jusqu'à atteindre 20 à 30 M.* (Astronomie)

Ex.14 : La proportion de cette eau qui a pu subsister jusqu'à maintenant est une question ouverte ; néanmoins, la Terre étant à ce stade très proche de la fin de son accrétion est un argument en faveur de l'idée que ceci correspond à la majorité de l'eau présente actuellement sur notre planète. (Astronomie)

*Accrétion* comme résultat :



Ex.15 : Adoptant cette valeur, et tenant compte de la géométrie du lac, l'aire d'accrétion couvrirait 53 % du lac (arrondi à 50 % dans ce qui va suivre). L'autre moitié du lac représente l'aire de fusion, ou pour le moins la zone où il n'y a pas d'accrétion. (Biologie)

Ex.16 : (...), ces corps de petite taille non différenciés et qui, de ce fait, sont sans doute des témoins privilégiés de l'état de la matière telle qu'elle existait, il y a 4,55 Ga environ, dans le disque d'accrétion, à une distance de l'ordre de 3 UA par rapport au Soleil naissant. (Chimie)

Ex.17 : la plaque nord-américaine est formée d'une mosaïque de morceaux de prismes d'accrétion, de terranes, plus anciens, représentant l'accrétion de la plaque américaine depuis 150 Ma. (Géologie)

Du point de vue de la fréquence, nous sommes donc en présence d'une discipline qui utilise *accrétion* à la fois comme processus et résultat (l'astronomie) et trois autres qui utilisent très majoritairement le sens d'un résultat (chimie, géologie, biologie).

Nous pouvons nous poser la question de savoir, dans ces cas de « restriction de sens », si nous sommes en présence d'un nouveau sens ou non. L'usage « résultat » est en effet un sens existant mais sa fréquence et sa quasi-exclusivité dans certains corpus amène à se demander si son sens est affecté.

Notons que l'utilisation de structures comme *durant/à la fin/pendant* sont assez connues pour identifier des déverbaux à valeur processive (par exemple, (Anscombe, 2007)). On aurait donc pu utiliser ces formes comme marqueurs (avec interprétation *a priori* donc) pour rechercher la présence de ces marqueurs systématiquement. Mais nous sommes dans des domaines spécialisés ; or, pour que la recherche de combinaisons de *accrétion* avec ces marqueurs soit envisagée, il aurait fallu avoir une hypothèse sur le fait qu'*accrétion* soit une nominalisation déverbale. Or, à ce moment de l'étude, nous ne savions pas qu'*accrétion* était un déverbal. Cet exemple montre ainsi que les corpus spécialisés demandent des stratégies d'analyse qui peuvent être assez différentes de celles mises en place dans des corpus non-spécialisés.

Un autre cas, présenté ci-dessous, illustre le fait que l'usage d'un terme entraîne des « déformations » du sens de ce terme au fil du temps. Comme Berrendonner et Béguelin le remarquaient pour la terminologie linguistique : « Les vocables dont sont composées les terminologies grammaticales ou linguistiques n'échappent pas au sort commun qui guette tous les lexèmes en usage ; leur signifié se déforme quand on s'en sert. Dès qu'ils sont mis en circulation, chacun se les approprie, et les plie au service d'intentions discursives pas toujours orthodoxes ; d'où des remodelages de leurs contenus, qui deviennent vite polysémiques et flous. S'agissant de vocabulaire scientifique, cette perpétuelle dérive des termes a de quoi dérouter (...) ». (Béguelin et Berrendonner, 2001 : 1).

#### Le cas de *fossile*

Prenons l'exemple de *fossile* dans le corpus d'exobiologie. Un nouveau sens de ce terme est en train de s'installer mais il y a, ici, une prise de conscience de ce nouveau sens qui « déforme » en quelque sorte le sens premier. Cette prise de conscience entraîne une certaine opposition à ce nouveau sens parce que, selon certains quelques uns, il déforme trop le sens premier. Ainsi en ce qui concerne le terme *fossile*, la définition est « toute trace de vie passée », sens que l'on retrouve dans les exemples suivants :

Ex.18 : Des fossiles de bactéries et des mattes bactériennes sont connus à toutes les époques géologiques (Westall, 1999 ; Westall et al., 2000).

Ex.19 : Les protéines ne seraient apparues que secondairement, les coenzymes et les ribozymes étant les traces fossiles des catalyseurs du passé.

Cependant, on retrouve des occurrences dans le corpus qui utilisent *fossiles vivants*, syntagme qui semble complètement contradictoire avec le sens donné ci-dessus :

Ex.20 : *Ainsi, les microsporidies, qui sont des eucaryotes simples (dépourvues de flagelles, d'appareil de Golgi, de mitochondries), ont souvent été qualifiées d'espèces primitives, de « fossiles vivants » de l'époque précédant l'endosymbiose mitochondriale.*

Ex.21 : *Par exemple, le coelacanth, l'exemple probablement le plus célèbre de « fossile vivant », a une morphologie globalement similaire à celle de certains fossiles vieux de quelques centaines de millions d'années mais il est très probable que son métabolisme est différent.*

On voit que les auteurs prennent une certaine précaution typographique en utilisant les guillemets parce qu'ils sont conscients que ce nouveau sens n'est pas encore totalement accepté par leur communauté. Certains contextes montrent même qu'il est complètement rejeté :

Ex.22 : *Malheureusement, la notion de « fossile vivant » est très discutable. En effet, toutes les espèces vivantes actuelles descendent d'un seul ancêtre commun, LUCA, et ont évolué durant exactement la même période de temps : elles sont donc toutes autant évoluées. Dire qu'une espèce actuelle est un fossile revient à considérer qu'elle a moins évolué que les autres. Or l'application de cette idée est parfaitement subjective.*

Comme le remarque Mortureux (1993) : » La ponctuation et la typographie (guillemets, parenthèses, caractères gras ou italiques) peuvent fonctionner comme marque, discrète et non univoque, d'activité métalinguistique, et notamment de reformulation. ». Jacobi (1985), dans son analyse du discours de vulgarisation scientifique, attribue aux guillemets un effet d'ambiguïté puisque parfois ceux-ci peuvent « proposer un raccourci familier d'une expression scientifique jugée complexe ».

Les guillemets sont un cas intéressant de la complémentarité entre interprétation *a priori* et *a posteriori*, c'est-à-dire avant ou après l'analyse du fonctionnement en corpus. En effet, ils montrent le possible changement de statut (d'indice à marqueur) qui peut s'effectuer pendant l'étude. Dans un premier temps, les guillemets jouent plutôt un rôle d'indice : le locuteur les utilise pour attirer l'attention sur un terme. Selon l'étude et le corpus choisi, cet indice sera interprété différemment (Rink et Tutin, 2007) ; ainsi, dans le corpus pluridisciplinaire, il est interprétable, dans la plupart des cas, comme un marqueur d'emprunt à une autre discipline.

Cette partie, décrivant des études de cas, nous a permis de mettre en évidence que, si les outils sont une aide précieuse, ils ne permettent pas de caractériser la nature des phénomènes sémantiques (restriction ou élargissement de sens, déformations, points de vue, précisions...). Par ailleurs, la dimension « vraiment » nouvelle des phénomènes langagiers est très difficile à repérer tant ces fonctionnements sont imbriqués avec des éléments extralinguistiques. L'analyse fine des contextes peut aider à repérer des différences de fonctionnements dont on fait l'hypothèse, compte tenu de la nature des corpus, qu'ils relèvent d'une nouveauté.

## Conclusion

La néologie sémantique est un phénomène difficilement repérable et caractérisable. Dans les corpus spécialisés, son étude passe nécessairement par la prise en compte d'éléments extralinguistiques et la confrontation avec l'avis des experts du domaine. Toutefois, une analyse préalable des corpus aide grandement à la mise au jour de changements dans les phénomènes

sémantiques. L'étude des corpus est désormais assistée par des outils qui proposent des résultats dont la prise en compte dans les analyses doit être minutieusement interrogée. Nous avons proposé une réflexion sur ce sujet qui s'articule autour du moment de l'interprétation : avant ou après la recherche en corpus. Dans le premier cas, nous parlons d'indices, dans le second cas, de marqueurs.

En effet, nous faisons l'hypothèse que certaines caractéristiques, de fréquence ou distributionnelles, ont un sens avant leur recherche en corpus. Ainsi, on peut penser que certaines structures langagières peuvent être associées de manière stable (c'est-à-dire hors corpus) à une interprétation, elle-même liée à l'objectif de l'étude. De la même façon, une variation importante de fréquence est le signe d'un changement d'usage majeur qui, dans le cas de la diachronique est interprétable d'emblée. Ces fonctionnements sont considérés comme des marqueurs. Inversement, l'étude de la distribution d'un terme ne relève pas d'une interprétation (d'une attente) *a priori*. On parlera alors plutôt d'indice.

Mais le plus souvent, hypothèse et confrontation aux données se succèdent au cours de l'analyse et s'alimentent mutuellement pour permettre la construction d'une interprétation sémantique qui sera soumise aux experts.

En corpus scientifique, l'interprétation est particulièrement délicate. En effet, les enjeux sémantiques ne concernent pas que le seul aspect langagier mais affecte aussi la position d'une discipline par rapport à d'autres disciplines et sa propre évolution.

Nous avons insisté sur deux points. D'une part, l'utilisation d'outils est très utile mais doit être réfléchie dans le cas d'une interprétation sémantique. Les indices fournis sont peu nombreux quant à leur nature (fréquence, variation de forme, distribution) et doivent être intégrés en fonction de l'objectif de l'étude et du moment de l'analyse. D'autre part, l'interprétation relève d'une construction qui, particulièrement dans le cas de domaines scientifiques, doit être mise en œuvre conjointement avec les experts. *In fine*, l'important n'est pas tant que l'analyste déclare que tel ou tel phénomène relève de la néologie sémantique mais bien que les experts prennent conscience qu'un phénomène de nouveauté repérable par des indices ou marqueurs linguistiques s'est mis en place, souvent à leur insu, et que c'est un élément important dans leur réflexion de scientifiques.

**Anne Condamines (1)**

**Nathalie Dehaut (1) et (2)**

**Aurélien Picton (3)**

(1) CLLE-ERSS, UMR 5263, CNRS et Université Toulouse Le Mirail

(2) CNES

(3) TIM-ISSCO, Faculté de Traduction et Interprétation, Uni-Mail, Genève

## Références

- Adam J.-M., 1999 : *Linguistique textuelle. Des genres de discours aux textes*. Paris : Nathan.
- Anscombe J.-C., 2007. « Les indicateurs aspectuels de déroulement processif : 'en cours de', 'en passe de', 'en train de', 'en voie de' ». *Cahiers de lexicologie* 90, p. 41-74.
- Assal A., 1992 : *Vocabulaire des biotechnologies : une approche descriptive*. Thèse de doctorat de sciences du langage. Université de Rouen.
- Bastuji J., 1974 : *Aspects de la néologie sémantique*. *Langages*, n°36, p. 6-19.
- Beacco, J. C. et Moirand, S., 1995 : *Les enjeux des discours spécialisés*. Paris : Presses de la Sorbonne Nouvelle.
- Béguelin M.L. et Berrendonner A., 2001 : « Circulation des termes et dérive terminologique ». In Colombat B. et Savelli M. (éds). *Métalangage et terminologie linguistique*. Leuven-Paris-Sterling : Peeters, p. 29-41.
- Bertels, A., 2006 : *La polysémie du vocabulaire technique. Une étude quantitative*, Thèse de doctorat non publiée, Université de Leuven, Belgique.
- Biber D., Johansson S., Leech G., Conrad S., Finegan E., 2000 : *Grammar of Spoken and Written English*. London : Longman.
- Bourigault, D., 2007 : *Un analyseur syntaxique opérationnel*. Mémoire d'Habilitation à Diriger les Recherches. Université Toulouse2 Le Mirail.  
w3.erss.univ-tlse2.fr/membres/bourigault/hdr.html
- Bourigault D. et Galy E., 2005 : « Analyse distributionnelle de corpus de langue générale et synonymie ». In *4e Journées de la linguistique de corpus*, Lorient, France, p. 163-174.
- Bowker, L. et Pearson, J., 2002 : *Working with Specialized Language : a Practical Guide to Using Corpora*. Routledge, London/New York.
- Condamines A., 2002 : « Corpus Analysis and Conceptual Relation Patterns ». *Terminology*, volume 8 number 1. p. 141-162.
- Condamines, 2005 : « Sémantique et Corpus, quelles rencontres possibles ? », In A. Condamines (ed.) : *Sémantique et Corpus*. Londres : Hermès.
- Condamines, A. et Rebeyrolle, J., 1997 : « Point de vue en langue spécialisée. », In *META*. XLII (1), p. 174-184.
- Condamines A. et Dehaut N., 2011 : « Mise en œuvre des méthodes de la linguistique de corpus pour étudier les termes en situation d'innovation disciplinaire : le cas de l'exobiologie », In *META*, LVI(2), p. 266-283.
- Fourez, G. : *La construction des sciences*, De Boeck Université, 2001.
- Gargaud M., Despois D. et Parisot J-P (eds), 2001 : *L'Environnement de la Terre Primitive*, Collection : *L'origine de la Vie sur Terre et dans l'Univers*, Presses Universitaires de Bordeaux.
- Guilbert L., 1975 : *La créativité lexicale*. Collection « Langue et Langage », Paris : Larousse Université.
- Jacobi, D., 1985 : « Sémiotique du discours de vulgarisation scientifique », In *Semen*, 2, <http://semen.revues.org/4291>.
- Koestler, A., 1965 : *Le cri d'Archimède*, Paris : Calmann-Lévy.
- Leech, G., 1992 : « Corpora and theories of linguistic performance », In J. Svartvik (ed) : *Directions in corpus linguistics*. Mouton de Gruyter : Berlin/New-York, p. 105-122.
- Mason, Z.J., 2004 : « CorMet : A Computational, Corpus-Based Conventional Metaphor Extraction System », In *Computational Linguistics*, 30 (1).
- Mortureux, M.-F., 1993 : « Paradigmes désignationnels », In *Semen*, 8, <http://semen.revues.org/document4132.html>.
- Nyckees V., 1997 : « Pour une archéologie du sens figuré », In *Langue Française*, 113, p. 49-65.

- Oliveira I., 2009 : *Nature et fonction de la métaphore en science, l'exemple de la cardiologie*. Paris : L'Harmattan.
- Paty, M., 1999 : « La création scientifique selon Poincaré et Einstein ». In Serfati, Michel (éd.), *La recherche de la vérité*, Coll. L'écriture des Mathématiques, ACLE. Paris : Editions du Kangourou, p. 241-280.
- Perlerin V., Ferrari S. et Beust P., 2005 : « Métaphores et dynamique sémique : expériences sur corpus ». In Williams G. (dir), *La Linguistique de Corpus*. Rennes : Presses Universitaires de Rennes, p. 323-336.
- Picton A., 2009 : *Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial*. Thèse de doctorat en Sciences du Langage, Université Toulouse 2. <http://tel.archives-ouvertes.fr/tel-00429061/fr/>.
- Rastier F., 2001 : *Arts et Sciences du texte*. Paris : Presses Universitaires de France, formes sémiotiques.
- Rayson, P.E., 2002 : *Matrix: a statistical method and software tool for linguistic analysis through corpus comparison*. Thèse de doctorat non publiée. UK : Lancaster University.
- Rinck F. et Agnès Tutin A., 2008 : « Annoter la polyphonie dans les textes : le cas des passages entre guillemets », *Corpus*, n°6. <http://corpus.revues.org/index1102.html>
- Sablayrolles, J.-F., (ed) 2008 : *Néologie et terminologie dans les dictionnaires*, Paris : Honoré Champion Editeur (collection « Lexica – Mots et Dictionnaires »).
- Schlanger J., 1991 : « La pensée inventive » In : Stengers I., Schlanger J. *Les concepts scientifiques : invention et pouvoir*, Paris : Gallimard, 1991.
- Slodzian M., 2000 : « L'émergence d'une terminologie textuelle et le retour du sens ». H.Béjoint et P.Thoiron (eds) : *Le sens en terminologie*. Lyon : Presses Universitaires de Lyon. p. 61-85.
- Teubert, W., 2002 : « Corpus Linguistics and Lexicography ». In *International Journal of Corpus Linguistics*, Vol. 6 (numéro special), p. 125-154.
- Tognini-Bonelli, E., 2001 : *Corpus Linguistics at Work*. Amsterdam/Philadelphia : John Benjamins Publishing Company.
- Ungureanu L., 2003 : *L'interpénétration langue générale-langue spécialisée dans le discours d'internet*, Paris : Connaissances et savoirs, Thèse de doctorat en Sciences du langage : Paris13. Thèse de doctorat en Sciences du langage : Université technique de Moldavie : 2003.
- Wüster, E., 1981 : « L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses ». In : G. Rondeau et H. Felber, dir. *Textes choisis de terminologie*. Québec : GIRSTERM, Université de Laval, p. 55-108.